Two Strongly Truthful Mechanisms for Three Heterogeneous Agents Answering One Question

Grant Schoenebeck, University of Michigan
Fang-Yi Yu, Harvard University

Motivation Questions

How can we design mechanisms to collect agents' truthful reports without verification?

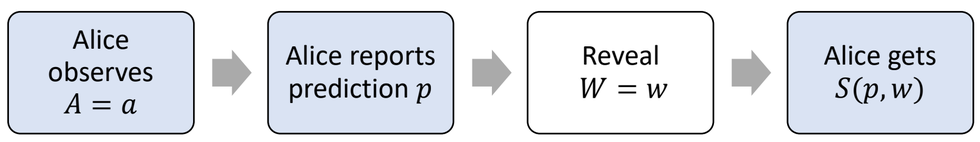
- three agents
- single item report
- asymmetric priors
- prior free

What is special about Bayesian Truth Serum?

- information score
- Log scoring rule

Proper scoring rule with Verification

Proper scoring rule $S(p,w)\in\mathbb{R}$



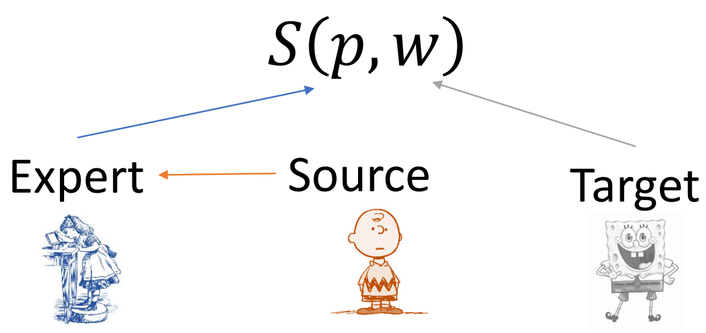
Log scoring rule $LS(p,w)=\log p(w)$

- Proper: truth-telling $P_{W|A}$ maximizes $LS(p,w)$
- Shannon mutual information:
$$\mathbb{E}_{W,A}\left[LS\left(P_{W|A},W\right)\right]=MI(W;A)$$
- Chain rule: If $p_i=P_{W|A_1\dots A_i}$ for all i
$$\mathbb{E}[LS(p_i,W)-LS(p_{i-1},W)]=MI(W;A_i\mid A_1,\dots,A_{i-1})$$

Proper Scoring Rule without Verification

Agents can play one of three roles

- Target reports his signal and is predicted.
- Expert makes prediction the target's signal.
- Source provides information to the expert.



Differential Peer Prediction Mechanisms

Theorem. DPP Mechanism is strongly truthful:

1. The truth-telling strategy profile is a strict Bayesian Nash Equilibrium,
2. The ex-ante agent welfare in the truth-telling strategy profile is strictly better than all non-permutation strategy profiles,

if the common prior is second order stochastic relevant on a finite set with full support.

DPP mechanism

1. Bob and Charlie report their signals B and C
2. Set one as Target and the other as the Source randomly. Set Alice as the expert.
3. Alice reports her initial prediction Q on **Target's** signal.
4. Alice learns **Source's** signal.
5. Alice updates her improved prediction Q^+ on **Target's** signal.

Two ideas for payments

- Source incentive: Using Source's (Charlie's) report to improve prediction $Q\to Q^+$
- Target incentive: Using Target's (Bob's) report as the ground truth

Payment	Source DPP	Target DPP
Alice (Expert)	$LS(Q^+,B)+LS(Q,B)$	
Bob (Target)	0	$LS(Q^+,B)$ $-LS(Q,B)$
Charlie (Source)	$LS(Q^+,B)$ $-3LS(Q,B)$	$-2LS(Q,B)$
Total	$2(LS(Q^+,B)-LS(Q,B))$ $=2MI(C;B\mid A)$	

Log scoring rule reversed

- Truthfulness the second argument?
- Reversed Log scoring rule
$$R(w)=LS(Q^+,w)-LS(Q,w)$$
- Symmetry between Target and Source
$$\mathbb{E}[LS(Q^+,B)-LS(Q,B)]=MI(C;B\mid A)$$

Connection to BTS and other mechanisms

Bayesian Truth Serum

- Each i agent reports her signal x_i and prediction p_i on other's signal.
- Prediction score: measure the quality of p_i
- Information score: $LS(Q^+,x_i)-LS(Q,x_i)$

  Aggregated prediction One agent's prediction
- Target-incentive mechanism

Source-incentive mechanism

- Robust BTS,
- Shadowing mechanism,
- Knowledge-Free Peer Prediction